

The Role of Computational Outage in Dense Cloud-Based Centralized Radio Access Networks

Matthew C. Valenti*, Salvatore Talarico*, and Peter Rost†

*West Virginia University, Morgantown, WV, USA.

†NEC Laboratories Europe, Heidelberg, Germany

Abstract—Centralized radio access network architectures consolidate the baseband operation towards a cloud-based platform, thereby allowing for efficient utilization of computing assets, effective inter-cell coordination, and exploitation of global channel state information. This paper considers the interplay between computational efficiency and data throughput that is fundamental to centralized RAN. It introduces the concept of computational outage in mobile networks, and applies it to the analysis of complexity constrained dense centralized RAN networks. The framework is applied to single-cell and multi-cell scenarios using parameters drawn from the LTE standard. It is found that in computationally limited networks, the effective throughput can be improved by using a computationally aware policy for selecting the modulation and coding scheme, which sacrifices spectral efficiency in order to reduce the computational outage probability. When signals of multiple base stations are processed centrally, a computational diversity benefit emerges, and the benefit grows with increasing user density.

Index Terms—Computational complexity, computational outage, turbo-decoding, mobile networks, 3GPP LTE

I. INTRODUCTION

In an information society as we have it today, mobile voice and data communication is a commodity service that is supposed to be available everywhere at any time. Accordingly, novel technologies to improve system capacity and quality of service are proposed and discussed in the context of next generation mobile networks. However, the focus of this discussion has to date focused primarily on the system performance, while the increased demand for computational resources has received relatively little attention.

A. Computational requirements in mobile networks

Mobile networks are hard real-time systems with tight timing and protocol constraints. These constraints are described by mobile communications standards such as 3GPP LTE [1] and must not be violated. In order to fulfill these constraints, a pre-defined amount of computational resources must be provided such that downlink and uplink processing can be performed within a given time interval. For instance, the uplink modulation and coding scheme (MCS) determines the number of information bits that must be processed per subframe.

The research leading to these results has received partly funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 317941 (www.ict-ijoin.eu). The authors would like to acknowledge the contributions of their colleagues in iJOIN, although the views expressed are those of the authors and do not necessarily represent the project.

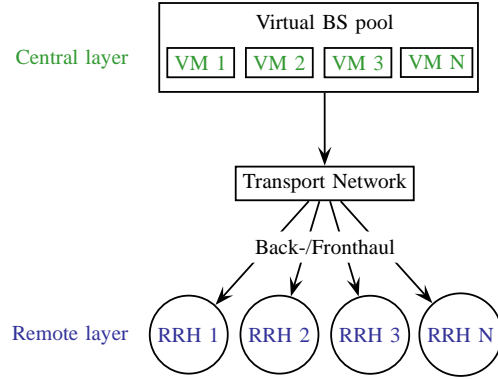


Fig. 1. Typical cloud RAN architecture

The current trend is towards the densification of mobile networks, where smaller cells are deployed. Compared to traditional macro-cellular base stations, small-cell base stations will be characterized by fewer antennas, different antenna and wave propagation patterns, lower transmission power, and limited computational resources. The latter characteristic is a consequence of economic constraints and the need for compact radio access points. Limiting the amount of computational resources may lead to *computational* outage rather than channel outage. Such computational outages occur whenever a decoding failure occurs due to the violation of a timing constraint rather than due to insufficient channel conditions.

Computational outage leads to a waste of spectral resources and loss of throughput, in much the same way as a channel outage. It is feasible that for some deployments, computational outage could be the dominant form of outage, and as such, will influence the design of the scheduler and certain aspects of the network architecture. The use of a *computationally aware* scheduler could yield great benefits when used in a network with computational constraints. Furthermore, the pooling of computational assets in a *cloud* could yield a *computational diversity* benefit, whereby centralized processing elastically focuses resources on those cells that have high momentary computing requirements.

B. Related work

The deployment of very dense networks requires novel technologies to allow for improved interference coordination and to improve the utilization of networks. Already today, net-

works are underutilized at less than 40 % maximum load. One technology to improve interference coordination and network utilization is centralized radio access network (RAN) [2]–[4]. Centralized RAN achieves this improvement by centralizing all baseband processing to a central entity. Currently deployed centralized RAN solutions exploit utilization improvements on a base station level; i.e., each baseband unit in a pool of units may be statically assigned to a particular remote radio head (RRH).

Using cloud computing as a platform for centralized RAN allows for the use of virtualization technologies that offer flexible provisioning of computational resources. Consider Fig. 1, where the RAN is divided into a remote and central layer. Depending on the backhaul quality, parts of the radio protocol stack are centralized and executed within a virtual base station (BS) pool [4]. Each virtual machine (VM) within this virtual BS pool may represent a radio access point (RAP) or cluster of RAPs, where each RAP may be a fully centralized base station (RRH) or just a partially centralized base station. Based on the actual traffic and resulting computational demand, each VM is provisioned with resources that can be elastically adjusted. One such split between remote and central layer could be to perform operations common to all users of a BS (e.g. FFT) at the RAP and to perform user-specific operations such as forward error correction (FEC) at the central layer [4], [5]. FEC consumes a major part of the computational resources, particularly in the uplink [6]. Hence, centralizing FEC (and all functionality above) allows for centralizing a large part of the base station complexity and provides opportunities to exploit the multi-user computational diversity.

However, in order to exploit such a cloud-computing platform efficiently, new technologies are needed to predict, monitor, and control the computational requirements. This requires the joint operation and optimization of the mobile network’s communication and computation systems. Recently, the computational complexity of mobile communication has received increasing attention, in particular for short-distance communication where power consumption due to transmitter and receiver processing may be on the order of the transmission power. In [7], Grover *et al.* showed that decoding complexity would scale at moderate to high signal-to-noise ratio (SNR) with $O(\log(x)^{-1})$ in the ratio of necessary SNR to achieve capacity and necessary SNR to achieve the chosen data rate. Hence, to get closer to capacity, the required complexity increases super-linearly.

C. Contribution and Outline

When small cells are provided with limited computational resources or multiple small cells share the (virtualized) computational resources, computational outage becomes increasingly important. However, determining computational outage is very difficult as it requires a detailed complexity model (similar to channel outage), and computational outage depends on multiple parameters such as SNR, block-length, modulation-and-coding-scheme, and decoder implementation.

In this paper, we analyze the likelihood of computational outage in a mobile communication network. More specifically, we first derive a model for computational outage and show how it relates to channel outage. In order to evaluate the actual impact on the throughput performance, this model is applied to a block Rayleigh fading channel and a network with co-channel interference. Local processing of the uplink signal is compared with central processing, and the computational diversity benefit of central processing is noted. We further introduce a computationally aware MCS selection policy that reduces the computational complexity requirements at the cost of slightly decreased spectral efficiency. However, in complexity constrained deployments, the selection policy provides higher effective throughput.

The paper is structured as follows: Section II introduces the notion of computational outage and relates it to channel outage, Section III analyzes the outage behavior in the case of a single cell, Section IV extends the analysis to a multi-cell environment where co-channel interference has to be considered. The paper is concluded in Section V.

II. CHANNEL AND COMPUTATIONAL OUTAGE

Consider a *cloud group* containing N_{cloud} base stations or RAPs whose signals are jointly processed in a virtual BS pool. Let Y_i indicate the i^{th} RAP and its location within the network. When a mobile in the cell served by Y_i transmits, a transport block (TB) is received at Y_i with a signal-to-interference-and-noise ratio (SINR) equal to γ_i . The SINR is assumed to be fixed for the duration of the subframe, which is true if the desired and all the interfering signals are subject to block fading¹, and hence the channel is AWGN with SNR γ_i .

When a TB is received with SINR γ , there is some probability that it will not be correctly decoded. Ideally, it will always be correctly decoded if γ is above some threshold and will be incorrectly decoded if it is below that threshold (a “brick wall” error-rate curve). However, this behavior requires infinitely long codes. In practice, a finite-length code, such as a turbo code, must be used, and its error-rate curve will not instantaneously drop to zero. To model this effect, we define a *random* error-indicator function $E(\gamma)$ which returns a 1 if a particular TB received with SINR γ fails to be successfully decoded and a 0 if it is correctly decoded. A *channel* outage occurs whenever $E(\gamma) = 1$. It follows that the channel outage probability is

$$\epsilon_{\text{channel}} = \mathbb{P}[E(\gamma) = 1] = \mathbb{E}[E(\gamma)]. \quad (1)$$

As the TB format and the number of turbo decoder iterations will vary from one transmission to the next, we furthermore define a random function $\mathcal{C}(\gamma)$ to be the *computational effort* required to process a particular TB received with SINR γ . Unless the link layer decides to drop a TB to prevent a computational overflow, every received TB must be processed. Thus, even if a channel outage occurs, $\mathcal{C}(\gamma)$ will generally

¹In the present paper, we use block fading as it eases the exposition. Fast correlated fading is for future study.

be nonzero. However, using an early-stopping criteria in the decoder may allow $\mathcal{C}(\gamma)$ to be less than its maximum value when a channel outage occurs.

The virtual BS pool will be provisioned with a limited amount of computational resources \mathcal{C}_{\max} per RAP. The units of these resources are the same as those associated with the computational effort $\mathcal{C}(\gamma)$. If the total required computational effort exceeds $N_{\text{cloud}} \cdot \mathcal{C}_{\max}$ before a decoding deadline is reached, a *computational* outage will occur. It follows that the computational outage probability for the cloud group is

$$\epsilon_{\text{comp}} = \mathbb{P} \left[\sum_{i=1}^{N_{\text{cloud}}} \mathcal{C}(\gamma_i) > N_{\text{cloud}} \cdot \mathcal{C}_{\max} \right]. \quad (2)$$

When a computational outage occurs, not all of the TBs in the cloud group are necessarily lost. If the processing is suitably scheduled, it is possible for some of the TBs to be correctly decoded, though there will not be enough resources to correctly decode all of the TBs in the group. The actual set of TBs that are lost depends highly on the deployed task scheduling algorithm. Thus, a TB can be lost due to a channel outage or a computational outage. Since even those TBs that are in a channel outage must be processed, these two kinds of outages are not mutually exclusive. Indeed, it is possible for a TB in a channel outage to trigger a computational outage as the baseband processor attempts to decode it. The *overall* outage probability ϵ , or simply *outage* probability, is the probability that a TB is lost due to either kind of outage.

Due to fluctuations in channel quality and traffic load, the SINRs at the different RAPs in the cloud group will generally be quite different, and as a consequence the offered computational loads $\mathcal{C}(\gamma_i), i \in \{1, \dots, N_{\text{cloud}}\}$, will vary from RAP to RAP. Centralizing the processing exploits this diversity of computational load, as computational assets can be diverted from RAPs with a currently low computational load to those with a high computational load. This *computational diversity* is a key benefit of cloud-based centralized radio access networks.

III. OUTAGE IN AN ISOLATED CELL

Consider a network with just one cell containing a single active user served by a single RAP. There is no interference from other cells, and the signal received by the RAP is processed all by itself; hence, $N_{\text{cloud}} = 1$. The statistics of $E(\gamma)$ and $\mathcal{C}(\gamma)$ are found through simulation. To provide meaningful results, the simulations used to determine these statistics are executed using parameters from the LTE standard.

A. Complexity-Throughput Tradeoffs in LTE

LTE features adaptive modulation and coding using turbo codes and hybrid ARQ. The RAP (called an *eNodeB*) commands the mobile (called a *UE*) to transmit using one of 27 distinct MCSs. Each MCS is identified by an MCS index, $I_{\text{mcs}} = \{0, \dots, 26\}$, and is characterized by a different combination of code rate and modulation format [8], [9]. Three kinds of modulation are used: QPSK ($0 \leq I_{\text{mcs}} \leq 10$), 16-QAM ($11 \leq I_{\text{mcs}} \leq 20$), and 64-QAM ($21 \leq I_{\text{mcs}} \leq 26$). When a TB is larger than 6144 bits, it is segmented into multiple

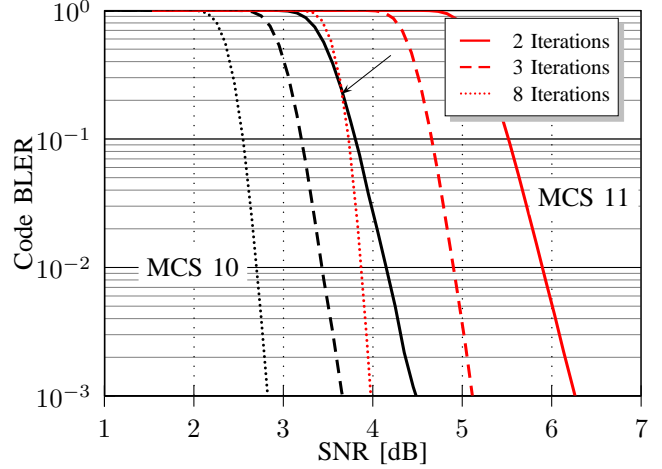


Fig. 2. Code block error rate as a function of SNR for MCS 10 and MCS 11 after 2, 3, and 8 decoder iterations. The arrow shows where the CBLER of MCS 10 after 2 iterations is the same as that of MCS 11 after 8 iterations.

code blocks (CBs). Each CB is separately turbo encoded, and all C CBs in the TB must be correctly decoded for the TB to be correct. Let ϵ_{cb} be the probability that a CB is in a channel outage. It follows that $\epsilon_{\text{channel}} = 1 - (1 - \epsilon_{\text{cb}})^C$ is the probability that the TB is in a channel outage.

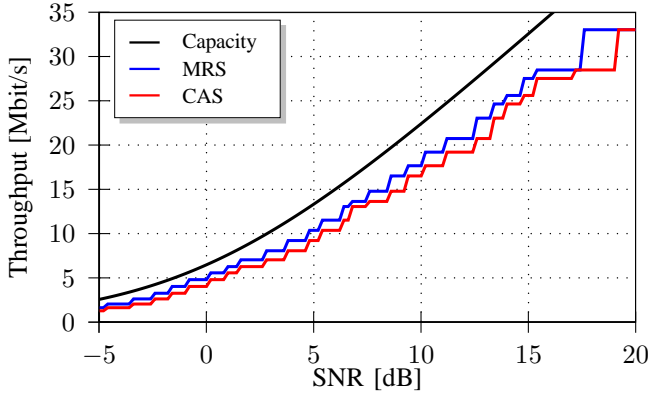
Turbo decoding is a computationally demanding task. However, because decoding is iterative and stops when a CRC check is satisfied, the amount of computational effort devoted to turbo decoding is highly variable. To handle this challenge, a centralized RAN deployment could virtualize its processing resources and flexibly assign turbo decoding tasks to available computing resources. Ultimately, it is the turbo decoding that dominates the computational outage behavior because it consumes a major part of baseband processing resources [6] and a TB will be lost if the turbo-decoding processes require more than the available computational resources.

The computational effort required to decode a turbo code is linear in the number of iterations and in the number of (information) bits. Thus a reasonable metric for computational effort is the *bit-iteration*, defined as follows. Let I_r and K_r be the number of executed iterations and the number of information bits associated with the r^{th} CB of the TB, respectively. In units of bit-iterations, the computational effort associated with the TB is

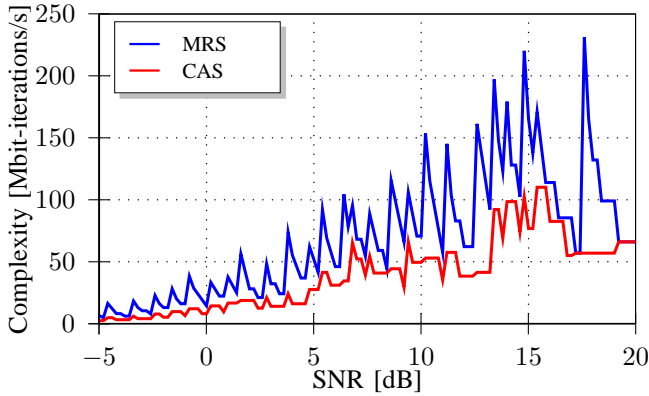
$$\mathcal{C}(\gamma) = \sum_{r=1}^C K_r I_r \quad (3)$$

where the dependence of K_r and I_r on γ is implicit.

The iterative nature of turbo decoding allows complexity to be traded off for transmission rate. This is illustrated in Fig. 2, which shows the code block error rate (CBLER) ϵ_{cb} as a function of SNR (expressed as the ratio of the energy-per-symbol to noise power, \mathcal{E}_s/N_0) after 2, 3, and 8 iterations for two MCSs: $I_{\text{mcs}} = 10$ and 11. The results in Fig. 2 and the remaining examples in this paper assume that the UE is allocated 45 resource blocks (RBs), which is the maximum



(a) Raw throughput



(b) Computational effort

Fig. 3. Raw throughput and computational effort as a function of the SNR for two MCS selection schemes: computationally aware selection (CAS) and max-rate selection (MRS).

allocation in a 10 MHz deployment when between three and five RBs are reserved for the physical uplink control channel [8], [9]. At $\epsilon_{cb} = 2.5 \times 10^{-1}$, the performance of MCS 10 with 2 iterations is the same as the performance of MCS 11 with 8 iterations. Thus, by backing off from MCS 11 to 10, the number of required iterations is cut by a factor of 4, on average. The number of information bits in the TB that must be decoded is also reduced, in this case from 9216 to 8064, so the required number of bit-iterations is even further reduced. However, this reduction in complexity comes at a cost, as the spectral efficiency of MCS 10 is lower than that of MCS 11.

The MCS is selected to satisfy a channel outage constraint $\hat{\epsilon}$. In this paper, we use $\hat{\epsilon} = 0.1$, which is a typical value for an LTE network. When complexity is not a concern, the selected MCS is the highest one that satisfies $\epsilon_{channel} \leq \hat{\epsilon}$ after a large number of decoder iterations. However, when complexity is a concern, the MCS should be selected such that the outage constraint is met after a specific number of decoder iterations. Define $R_i(\gamma)$ to be the maximum rate for which $\epsilon_{channel} \leq \hat{\epsilon}$ after the i^{th} decoder iteration, where the maximization is over the set of MCSs. The function $R_i(\gamma)$ determines the MCS-selection scheme. When complexity is

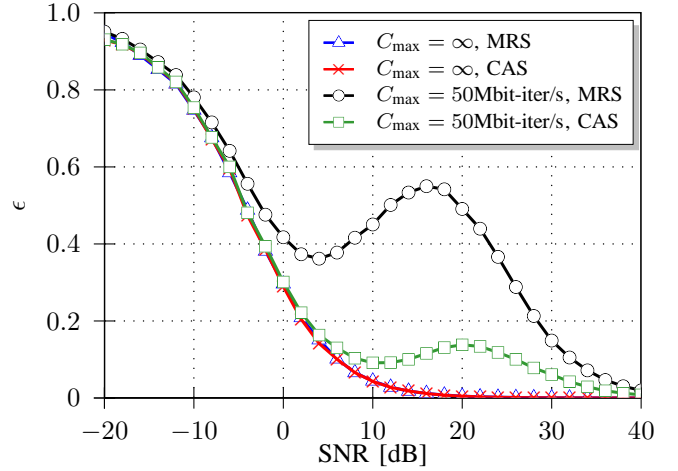


Fig. 4. Outage probability as a function of average SNR in the presence of Rayleigh fading, both with and without a complexity constraint.

not a concern, $R_i(\gamma)$ with a large value of i could be used, e.g. $i = 8$. In a complexity constrained deployment, using $R_i(\gamma)$ with a smaller value of i could be used, where a sensible value might be $i = 2$. In the following, we refer to the *max-rate selection* (MRS) policy as the one that selects the MCS that achieves $R_8(\gamma)$, while we refer to the *computationally aware selection* (CAS) policy as the one that selects the MCS that achieves $R_2(\gamma)$. The minimum value of $\Delta\gamma$ for which $R_2(\gamma + \Delta\gamma) = R_8(\gamma)$ is an SNR margin required when CAS is used instead of MRS.

The rate-complexity tradeoff is illustrated in Fig. 3. Fig. 3(a) shows the raw throughput of the MRS and CAS schemes as a function of the SNR². For reference, the Shannon capacity limit is also shown. The reduction in rate when using CAS is evident from the figure. The average computational effort required to decode a TB as a function of SNR for the two MCS selection policies is shown in Fig. 3(b). As anticipated, using the complexity aware scheme reduces the average effort. The staircase appearance of Fig. 3(a) and peaky behavior and Fig. 3(b) are due to the use of a finite number of MCSs.

B. Impact of Rayleigh Fading

Now consider the complexity-throughput tradeoff in a fading channel. When the fading is Rayleigh, the instantaneous SNR γ is an exponential random variable with mean $\Gamma = \mathbb{E}[\gamma]$. The complexity-outage tradeoff is obtained through a simulation that works as follows. During each trial, the SNR γ is drawn from an exponential distribution. The MCS scheme for the given γ is determined according to the MCS selection policy. Each of the C CBs in the TB is marked as being in a channel outage with probability ϵ_{cb} , which can be precomputed. If any of the CB are in outage, then the entire TB is declared to be in an outage. If the r^{th} CB is in an

²The *raw* throughput T_{raw} is simply the selected rate $R_i(\gamma)$ expressed in units of bits/second. This is in contrast with the *effective* throughput, which is the rate of *correct* transmission; i.e., $T_{eff} = (1 - \epsilon)T_{raw}$.

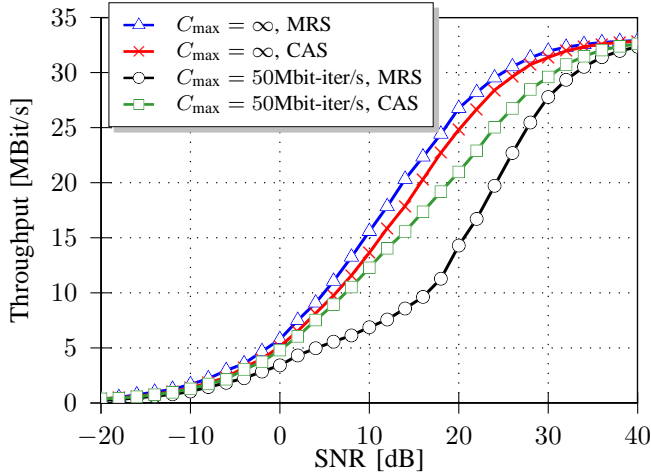


Fig. 5. Effective throughput as a function of average SNR in the presence of Rayleigh fading, both with and without a complexity constraint.

outage, then $I_r = I_{\max}$, which is the maximum number of attempted iterations (typically 8). Otherwise, I_r is determined by drawing a random variable distributed according to the pdf of I_r , which can be precomputed by tracking the error-rate as a function of the number of iterations.

Fig. 4 shows the outage probability of both MCS-selection schemes in the presence of Rayleigh fading when there is no complexity constraint ($C_{\max} = \infty$) and when there is a complexity constraint of $C_{\max} = 50$ Mbit-iterations per second³. As can be seen, using the more conservative CAS scheme greatly reduces the outage probability when complexity is constrained. A non-monotonic behavior is observed in the complexity constrained outage curves. At low SNR, channel outage dominates, as the conservative MCSs that must be selected do not have particularly high computational requirements due to having few information bits per TB. However, as the average SNR increases, higher MCSs are selected more frequently and computational outage begins to dominate due to the additional computational burden associated with their larger payloads. However, at very high average SNR, the instantaneous SNR is often much higher than the selection threshold for the highest MCS, and the CBs for that MCS will usually be successfully decoded after just one or two iterations, which is not enough load to trigger a computational outage.

In a complexity constrained system, the advantage of a complexity aware MCS selection is a reduction in outage probability, but its disadvantage is a loss in the raw throughput. However, if the outage probability is sufficiently lower, the effective throughput may be higher despite the reduced raw throughput. This can be seen in Fig. 5, which shows the effective throughput as a function of average SNR for the same cases that were shown in Fig. 4. The loss in effective

³The complexity of a typical software turbo decoder requires between 100 and 1000 floating-point operations per second (FLOPS) for each bit-iteration, depending on the actual implementation [10]. For comparison, modern general-purpose processors support up to 50-200 billion FLOPS.

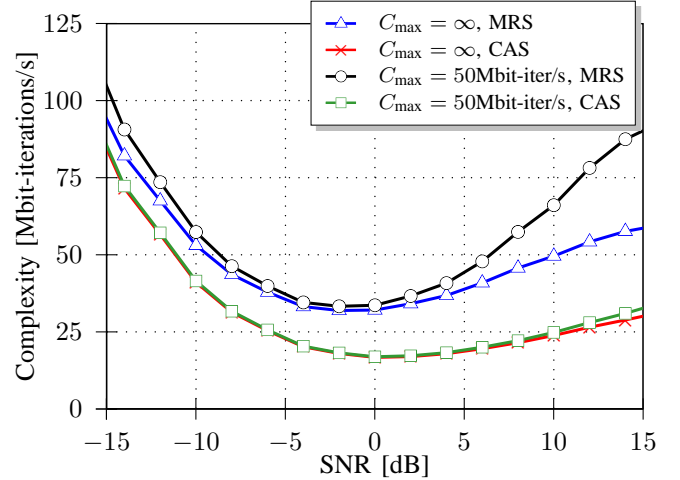


Fig. 6. Average complexity required in Rayleigh fading for a successfully decoded codeword as function of average SNR.

throughput due to the complexity constraint is evident in the figure, but much of this loss can be recovered by using complexity aware MCS selection; i.e., the throughput curve for CAS is up to 9 dB higher than the curve for MRS.

The average complexity required for successful decoding in Rayleigh fading is shown in Fig. 6. The MRS scheme always has a higher complexity load than the CAS scheme. Imposing a complexity constraint causes the average complexity for successful decoding to increase, due to the occurrence of computational outages, which results in wasted computations and a corresponding increase in average complexity. However, while this increase in complexity for the MRS is significant, especially at high SNR, the increase is barely noticeable for the CAS scheme. While there is still a chance of computational outage with the CAS scheme, the probability is too low to affect the average computational complexity.

IV. OUTAGE IN A NETWORK

Now consider a network with N_{total} base stations or RAPs, where N_{total} could be arbitrarily large. Unlike the single-cell scenario, there is co-channel interference on the uplink due to active mobiles in adjacent cells. For illustrative purposes, we consider the network with $N_{\text{total}} = 129$ shown in Fig. 7, which is a segment of an actual deployment by a major provider in the UK at 1800 MHz. We consider the processing of the $N_{\text{cloud}} = 8$ cells highlighted in yellow. Two options are considered, *cloud processing* (CP), where the signals from all N_{cloud} cells are processed in the cloud, and *local processing* (LP), where the signals at each RAP are locally processed.

We assume that the UEs are distributed according a Poisson point process (PPP) with intensity λ users per km^2 . The analysis of the cellular uplink with randomly distributed users has been recently considered in [11], [12]. While [11] assumes randomly placed base stations, the analysis in [12] allows for arbitrarily placed base stations, such as those shown in Fig. 7.

RAP Y_i serves a cell with area \mathcal{A}_i . The number of UEs in Y_i 's cell is a Poisson variable with mean $\lambda \mathcal{A}_i$. While the

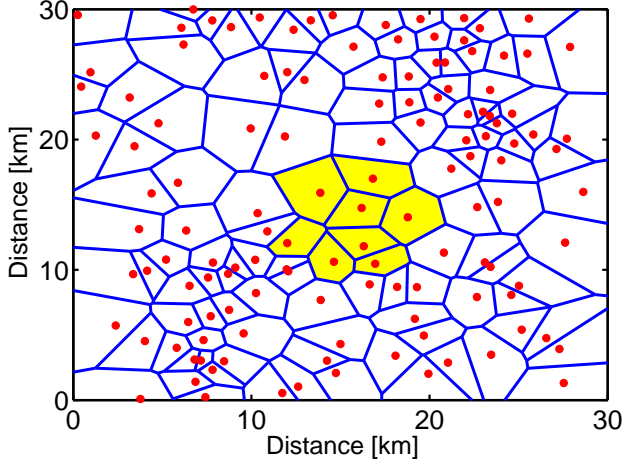


Fig. 7. Base station locations. The cloud group consists of the eight highlighted cells in the center of the diagram.

SC-FDMA uplink channel is partitioned into a plurality of RBs that could be allocated to different users, we assume that all of the available resources of a particular subframe are allocated to a single UE. Computationally, this is a worst-case scenario as there is no computational diversity within a given cell. Continuing the example from the previous section, we assume a 10 MHz system with 45 RBs allocated to each uplink user. When there is more than one user in Y_i 's cell, then one of the users is selected at random to transmit on the uplink. The likelihood of a transmitting user in Y_i 's cell is the complement of the void probability. Thus, there will be a UE transmitting to Y_i with probability $1 - \exp(-\lambda \mathcal{A}_i)$, and when there is a transmission, the location of the user is uniform within the cell. The location of the UE transmitting to Y_i is denoted by X_i .

The path loss from a mobile X to a base station Y is $|Y - X|^{-\alpha}$, where α is the *path-loss exponent*. X_i transmits its uplink signal with power P_i . The value of P_i is selected according to the fractional power-control policy:

$$P_i = P_0 |Y_i - X_i|^{s\alpha} \quad (4)$$

where P_0 is a reference power (typically taken to be the power received at unit distance from the transmitter) and $s, 0 \leq s \leq 1$, is the *compensation factor* for fractional power control. We assume $s = 0.1$, which is the value reported in [13] that maximizes the sum throughput. Higher values of s would improve fairness at the cost of throughput.

The fading power gain from X_i to Y_j is $g_{i,j}$, which is normalized to have unit mean. We again assume that $g_{i,j}$ is exponential (Rayleigh fading) and that the fading power gains remain fixed for the subframe, but vary from one subframe to the next (block fading). The SINR at Y_j is

$$\begin{aligned} \gamma_j &= \frac{P_j g_{j,j} |Y_j - X_j|^{-\alpha}}{W + \sum_{i \neq j} P_i g_{i,j} |Y_j - X_i|^{-\alpha}} \\ &= \frac{g_{j,j} |Y_j - X_j|^{\alpha(s-1)}}{\Gamma^{-1} + \sum_{i \neq j} g_{i,j} |Y_j - X_i|^{-\alpha} |Y_i - X_i|^{s\alpha}} \end{aligned} \quad (5)$$

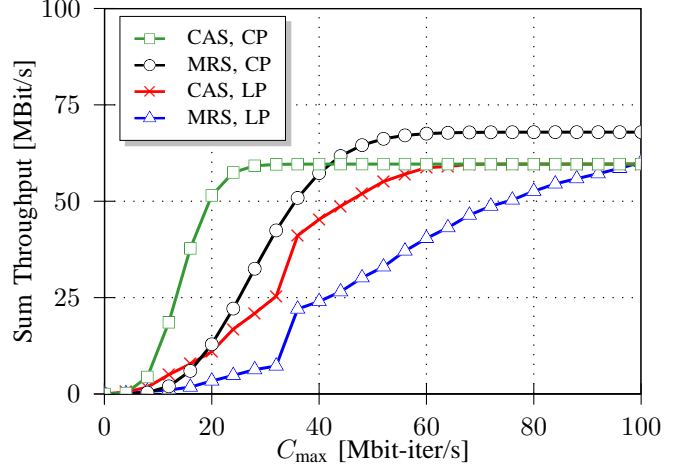


Fig. 8. Sum throughput as function of the per-RAP complexity constraint when $N_{\text{cloud}} = 8$, with local processing (LP) and with cloud processing (CP). Two MCS-selection schemes are considered: computationally aware selection (CAS) and max-rate selection (MRS).

where W is the noise power, $\Gamma = P_0/W$ is the SNR at unit distance, and the summation is over all the co-channel interferers.

As in the single-cell case described in Section III, the throughput is determined with the assistance of a Monte Carlo simulation. During each trial, which corresponds to a subframe, a mobile is placed at random in the i^{th} cell with probability $1 - \exp(-\lambda \mathcal{A}_i)$. Once the mobiles are placed, the fading coefficients $g_{i,j}$ are drawn from an exponential distribution and the SINR at each RAP in the cloud group is computed according to (5). The TBs that are in a channel outage are identified and the computational effort $C(\gamma_i)$ for each TB is determined as before.

Both LP and CP are considered. In the LP case, if $C(\gamma_i) > C_{\text{max}}$ for a given TB, then it is considered to be in an outage. For the CP case, if (2) is violated, then an outage occurs in at least one of the uplink TBs. Scheduling in the virtual BS pool is assumed to process the signals with low SINR first. This allows those TBs that are sent at a low MCS, and hence have a lower complexity requirement, to be first decoded, and the TBs that fail are the ones that were sent at the highest MCSs.

Fig. 8 shows the sum throughput, which is the sum of the effective throughputs in each cell of the cloud group, as a function of the per-RAP complexity constraint C_{max} with the two MCS-selection policies. The plot was generated assuming a UE density of $\lambda = 0.1$, path-loss exponent $\alpha = 3.7$, and average SNR $\Gamma = 20$ dB. As can be seen, CP is much more computationally efficient than LP. For the same complexity constraint, the throughput of CP is significantly higher than it is with LP. The computational diversity advantage of CP is evident from its steeper curve. Alternatively, the complexity constraint required to achieve a desired throughput is much lower for CP than it is for LP. These behaviors confirm the computational diversity benefits of cloud processing. Furthermore, there is a benefit to using the more conservative CAS

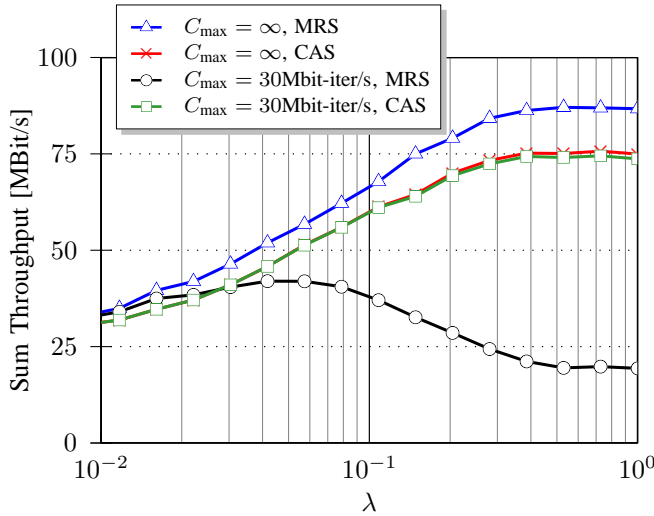


Fig. 9. Sum throughput with cloud processing as function of the density of UEs when $N_{\text{cloud}} = 8$. Two MCS-selection schemes are considered: computationally aware selection (CAS) and max-rate selection (MRS).

policy for the LP system. For the CP system, CAS is beneficial as long as $C_{\text{max}} \leq 40$ Mbit-iter/s; for systems with less of a complexity constraint MRS is better.

Fig. 9 shows the effect of mobile density λ on the sum throughput. Again, $\alpha = 3.7$, and the average SNR $\Gamma = 20$ dB. The figure shows results for CP, both with no complexity constraint ($C_{\text{max}} = \infty$) and with a constraint of $C_{\text{max}} = 30$ Mbit-iter/s. The two MCS selection policies are again considered. When there is no complexity constraint, the sum throughput rises with λ . In this case, the throughput of MRS is higher than CAS. However, when there is a complexity constraint, throughput falls with higher values λ when MRS is used. This is due to a high prevalence of computational outages in heavily loaded systems. However, if the more conservative CAS policy is used, performance is approximately the same as if there were no complexity constraint. This confirms the effectiveness of using a more conservative MCS-selection policy in dense networks.

V. CONCLUSIONS

Computational limitations in mobile wireless networks have a significant impact on the network performance in the moderate to high SNR regime. Such computational limitations are of particular interest to small-cell networks, where RAPs are computationally limited due to economic constraints. The computational constraints are also of fundamental importance to the design of centralized RAN architectures. With the framework introduced in this paper, a direct link between the employed computing technology and the communication technology can be drawn, which opens the opportunity to exploit new degrees of freedom within a mobile network.

The newly introduced concept of *computational outage* helps to quantify the complexity-throughput tradeoff in centralized RAN platforms. Such a quantification allows the computational requirements for centralized RAN platforms to be specified in a meaningful way. The analysis furthermore provides insight into the *computational diversity* that can be

achieved by cloud processing, and the corresponding increases in effective throughput.

This paper discussed a single-cell and a multi-cell scenario where inter-cell interference has a considerable impact on the network performance. It was found that in computationally constrained mobile networks, there is a benefit to intentionally selecting a lower MCS value, as its lower computational demand may actually increase the effective throughput due to a reduction in computational outage. The benefits of using cloud processing and using a conservative MCS selection policy become more pronounced as the user density increases.

The analysis in this paper has assumed block-fading channels. In practice, user mobility and time-variant channels will have a significant impact on the performance of a centralized RAN. The study of such systems is a key future direction for this research. It may be that the solution is to use an aggressive MCS-selection scheme for some users, and reserve the conservative MCS-selection scheme for only a small number of user so as to not overload the system. Such a strategy could be interpreted as a kind of “computational water-filling”, and could allow for reducing the ratio of peak-to-average computational effort in a centralized RAN, thereby increasing the utilization of the network infrastructure.

REFERENCES

- [1] The 3rd Generation Partnership Project. (2014, March) 3GPP. [Online]. Available: <http://www.3gpp.org>
- [2] X. Wang, Y. Huang, C. Cui, K. Chen, and M. Chen, “C-RAN: Evolution toward Green Radio Access Network,” *China Communications*, vol. 7, pp. 107–112, 2010.
- [3] H. Guan, T. Kolding, and P. Merz, “Discovery of Cloud-RAN,” in *Cloud-RAN Workshop*, Apr. 2010.
- [4] P. Rost, C. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, “Cloud technologies for flexible 5G radio access networks,” *IEEE Communications Magazine*, vol. 52, no. 5, May 2014.
- [5] U. Dötsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, “Quantitative analysis of split base station processing and determination of advantageous architectures for LTE,” *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105–128, 2013.
- [6] S. Bhaumik, S. P. Chandrasekhar, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, “CloudIQ: A Framework for Processing Base Stations in a Data Center,” in *18th Annual Inter. Conf. on Mobile Computing and Networking (MobiCom)*, Istanbul, Turkey, Aug. 2012.
- [7] P. Grover, K. A. Woyach, and A. Sahai, “Towards a communication-theoretic understanding of system-level power consumption,” *IEEE Journal on Selected Areas in Communications*, September 2011.
- [8] 3GPP, “Radio access network; evolved universal terrestrial radio access; (e-utra); physical channels and modulation (release 10),” 3GPP, Tech. Rep., Dec. 2012.
- [9] —, “Radio access network; evolved universal terrestrial radio access; (e-utra); physical channels and modulation,” 3GPP, Tech. Rep., Dec. 2012.
- [10] M. C. Valenti and J. Sun, “The UMTS turbo code and an efficient decoder implementation suitable for software defined radios,” *Int. J. Wireless Info. Networks*, vol. 8, pp. 203–216, Oct. 2001.
- [11] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, “Analytical modelling of uplink cellular networks,” *IEEE Transactions on Communications*, vol. 12, pp. 2669–2679, Jun. 2013.
- [12] D. Torrieri, M. C. Valenti, and S. Talarico, “An analysis of the DS-SS cellular uplink for arbitrary and constrained topologies,” *IEEE Transactions on Communications*, vol. 61, pp. 3318–3326, Aug. 2013.
- [13] M. Coupechoux and J. M. Kelif, “How to set the fractional power control compensation factor in LTE?” in *IEEE Sarnoff Symposium*, Princeton, NJ, May. 2011.